

PersID - eurooppalaista yhteistyötä pysyvien tunnusten parissa

Esa-Pekka Keskitalo
Kansalliskirjasto

Tämän artikkelin pysyvä osoite on
<http://urn.fi/URN:NBN:fi-fe201106081712>

Pysyvät tunnukset ovat viime aikoina olleet uudelleen virinneen mielenkiinnon kohteena. Erityisen ajankohtainen tunnusasia on tutkimusdatan suhteen, jonka järjestelmällinen ja pitkäaikainen säilyttäminen on paraikaa muotoutumassa. Myös valtion tietohallinnon kehittämisessä pysyvät tunnisteet on nostettu kehittämiskohteeksi.

Pysyviä tunnuksia on Tietolinjassa käsitelty aiemminkin. Viime numerossa (<http://urn.fi/URN:NBN:fi-fe201012143109>) oli Juha Hakalan ehdottomasti tutustumisen arvoinen ja perusteellinen selvitys pysyvistä tunnisteista. Numerossa 1/2007 (<http://urn.fi/URN:NBN:fi-fe20071298>) oli juttua URN-tunnuksista - vieläkin käypä esitys siitä, miten URN-tunnukset toimivat ja miksi niistä on hyötyä. Lyhyenä kertauksena:

- URN-tunnus on digitaalisen objektin pysyvä ja ainutkertainen tunnus. Objektin URN-tunnus ei koskaan muutu. Samaa URN-tunnusta ei ole koskaan kahdella toisistaan eroavalla digitaalisella objektilla.
- URN on siis erilainen tapa tunnistaa objekti ja viitata siihen kuin URL, joka on objektin osoite. Jos objektista on kaksi identtistä kopiota, niillä on eri URL mutta sama URN. Samassa URL-osoitteessa taas saattaa eri aikoina olla täysin erilaista informaatiota.
- Mitä pidempi aika kuluu, sitä mahdollisempaa on säilyttää URL-osoitteet muuttumattomina. Sellaiset ratkaisut kuin Cool URI toimivat toki yhden toimijan ja yhden järjestelmän sisällä, mutta URN-tunnuksen on tarkoitus toimia yli organisaatioiden ja teknologioiden muutosten

Sekavan kentän ongelmat

Kansalliskirjastokin on työskennellyt URN-tunnusten parissa melko yksin. Valmiita luontevia yhteistyöfoorumeita on puuttunut. Tämän johdosta eri maissa on syntynyt hieman erilaisia käytäntöjä ja noudatettu erilaisia periaatteita URN-tunnusten jakelussa. Suomi esimerkiksi taitaa olla ainoa maa, jossa annetaan myös ISBN-numeroon perustuvia URN-tunnuksia. Kansainvälisen yhteistyön puute on myös hidastanut URN-tunnusten käytön leviämistä, koska niiden luonne oikeastaan vaatisi kansainvälisiä palveluita.

Tunnuksien käytöstä kiinnostuneen on ollut hankala löytää tietoa pysyvien tunnusteiden kentästä sekä eri toimijoiden suunnitelmista ja tulevaisuuden näkymistä. Varsin näkyvästi oli esitetty myös mielipiteitä, että tunnukset eivät oikeastaan olisi tarpeellisia. (Tämän näkökulman torjuminen olisi oman artikkelinsa aihe. Tässä riittääköön sanoa, että monet kritisoijatkin myöntävät, että hyvin pitkällä aikavälillä tunnuksien hyödyllisyys lisääntyy.)



Euroopalle yhteinen suunta

PersID- hanke (<http://www.persid.org>) alkoi vuonna 2009. Siinä oli mukana kirjastoja, tutkimuslaitoksia ja muita organisaatioita Pohjoismaista, Alankomaista, Italiasta ja Saksasta. Yksi sen ensimmäisiä tehtäviä olikin selvittää pysyvien tunnisteiden käytön nykytilannetta.

Hankkeessa kiinnitettiin erityisesti huomiota pysyviin tunnuksiin osana pitkäaikaissäilytystä. Objektien luotettava tunnistus hyvin pitkiä aikoja vaatii oikeastaan välttämättä jonkinlaista tunnistejärjestelmää. Mutta asian voi nähdä myös toisin päin: ollakseen pysyvä, tunnuksen pitää pystyä viittaamaan todellisesti olemassa olevaan objektiin hyvin pitkiä aikoja. Näin ollen aineistojen pitkäaikaissäilytys varmistaa tunnusten toiminnan.

Eryteisesti nykytilanteessa tunnistettiin seuraavat kolme pysyvien tunnisteiden ongelmaa:

- **Luottamuksen puute.** Tunnusjärjestelmien moninaisuus ja ristiriitaisuus on johtanut siihen, ettei ole olemassa selkeitä periaatteita esimerkiksi sille, miten ja millaisille aineistoille tunnuksia annetaan. Järjestelmät vaikuttavat sekavilta ja siksi epäluotettavilta. Tietoa ja ohjeita tunnusten käytöstä on myös vaikea saada.
- **Epävarmuus pitkällä aikavälillä.** Pysyvien tunnusten järjestelmien ylläpitäjät eivät ole selkeästi sitoutuneet pysyvästi vastaamaan tunnusten toimivuudesta. Osa tunnusjärjestelmistä on taustoiltaan varsin kaupallisia, joten päätökset tulevaisuudesta eivät välttämättä synny esim. tutkimuksen tarpeita ajatellen.
- **Palvelujen puute.** Nykyinen tunnuskenttä on niin fragmentoitunut, ettei sen varaan voi rakentaa hyödyllisiä palveluita

PersID-hankkeessa ei ollut ensi sijassa kysymys teknologiasta. Tunnusjärjestelmien tekniikka ei ole mitenkään erityisen monimutkaista; suurimmat vaikeudet liittyvät pikemminkin selkeiden toimintamallien puutteeseen ja toimintatapojen erilaisuuteen. Hankkeessa etsittiinkin ennen kaikkea yhteistä, pysyvää sitoutumista siirtyä kohti suurempaa yhtenäisyyttä. Osallistujat ovatkin sitoutuneet jatkamaan aloitettua työtä hankkeen päättymisen jälkeenkin. Hankkeen aikana valmisteltiin luonnos yhteisesti noudatettavaksi politiikaksi (<http://persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-04p-8py> (<http://persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-04p-8py>), s. 23).

Hankkeen yhteydessä tutkittiin myös, millaisia kustannuksia eri maille koituu URN-järjestelmästä. Kustannukset näyttävät jakautuvan kaikkialla kolmeen jokseenkin yhtä suureen osaan: teknologian kustannuksiin kuten palvelimiin; asiakaspalvelun kustannuksiin sekä markkinoinnin, strategisen suunnittelun ja yhteistyöprojektien aiheuttamiin sekalaisiin kustannuksiin. On ilmeistä, että vaikka kustannukset kasvavat, kun tunnusten määrä kasvaa, niin tunnusten yksikköhinta pienenee merkittävästi. Suuret järjestelmät ovat siis suhteellisen tehokkaita. Jos ajattelemme pysyviä tunnuksia osana pitkäaikaissäilytyksen kokonaisuutta, niiden kustannusmerkitys on mitätön.

Suomen URN-resolverissa oli selvityksen aikoihin noin kuusi miljoonaa URN-tunnusta, joita kyseltiin noin 16 500 kertaa päivässä. Kaikkine, myös kehityskuluineen, URN-systeemi maksaa vuodessa reilusti alle 20 000 euroa.

NBN-numerot - joustavasti mutta kontrolloidusti

PersID-hankkeen yksi tärkeä periaate oli se, että kyseessä ei ole taistelu yhden tunnusjärjestelmän puolesta muita vastaan. Päinvastoin, hankkeen aikana on oltu tiiviisti yhteydessä etenkin DOI-tahoihin. Myös muut ovat nyt kiinnostuneet ajatuksesta kehittää tunnusten käyttöä yhdessä.

Eryteisesti pitkäaikaissäilytyksen tarpeista johtuen koettiin, että hankkeessa keskityttäisiin kuitenkin URN-tunnuksiin, ja etenkin NBN-numeroihin perustuviin URN-tunnuksiin. (NBN tulee sanoista "National Bibliographic Number". Se on kansalliskirjastojen hallinnoima tunnusjärjestelmä, josta voidaan rakentaa URN-tunnus. NBN-numeroita voidaan jaella hyvin joustavasti kaikenlaisille digitaalisille objekteilla.)

- URN:NBN-tunnuksia on käytetty hyvällä menestyksellä eri maissa
- Kansalliset ratkaisut on mahdollista yhdistää maailmanlaajuisiksi palveluksi
- Kansalliskirjastot ja muut vastaavat tutkimus- ja kulttuuriperintöorganisaatiot ovat valmiita ottamaan vastuuta URN:NBN-järjestelmän toimivuudesta ja pysyvyydestä.
- URN:NBN-tunnukset perustuvat avoimiin standardeihin; ne ovat erässä mielessä osa kansalliskirjastojen julkista palvelua.

URN:NBN-tunnusten avulla voidaan saavuttaa luotettuus, pitkän aikavälin pysyvyys ja maailmanlaajuisesti yhteentoimivien palvelujen verkosto. Tämän vuoksi hankkeessa tehtiin konkreettisesti kehitystyötä juuri URN:NBN-tunnuksiin liittyen.

Kohti globaalia palvelua

Käyttäjän kannalta olisi toivottavaa, että URN-tunnuksien - ja muiden pysyvien tunnuksien - tuki paranisi ainakin sikäli, että www-selaimet tunnistaisivat ne. Tällöin pelkkä URN-tunnus muodostuisi linkiksi, tällä hetkellähän täytyy erikseen rakentaa www-osoite, esim.

URN:NBN:fi-fe201105131534 ->
<http://urn.fi/URN:NBN:fi-fe201105131534>

Tähän tarkoitukseen on olemassa jo esim. [Firefoxiin lisäpalanen \(https://addons.mozilla.org/fi/firefox/addon/oid-resolver/\)](https://addons.mozilla.org/fi/firefox/addon/oid-resolver/). Ongelma on palveluiden rajallisuus: Suomen URN-resolveri resolvoi lähinnä suomalaisia tunnuksia, samoin Saksan saksalaisia jne. PersID hankkeessa tehtiin pohjatyötä sille, että olisi yksi, robusti "metaresolveri", joka ohjaisi kyselyt yhdestä palvelupisteestä aina oikeaan varsinaiseen resolveriin. Se siis tunnistaisi sen, että yllä olevassa esimerkissä on kyse suomalaisesta tunnuksesta ja ohjaisi kyselyn Suomeen. Metaresolverin arkkitehtuuri on kuvattu PersID-hankkeen loppuraportin osassa 4. Resolverista on jo olemassa prototyyppi, jonka lähdekoodi on saatavissa.

Tällainen resoluutiopalvelu ei hyödyttäisi ainoastaan URN-tunnuksia, vaan sama metaresoluutiopalvelu voisi aivan helposti käsitellä myös vaikkapa DOI-tunnuksia. Näin ollen olisi tarpeen tietää vain yksi resolverin osoite, ja tunnusten automaattinen käyttö erilaisissa tilanteissa olisi helppoa. Palvelun riskejä pitäisi tietenkin hajauttaa sopivilla järjestelyillä. Tämän palvelu rakentaminen jatkuu vielä.

URN-standardeja uudistetaan

Nykyiset URN-standardit ovat vuodelta 1997. Ei ole siis ihme, etteivät ne kaikilta osin vastaa nykyisiä tarpeita. Hankkeen rinnalla on, erityisesti Kansalliskirjaston Juha Hakalan aloitteesta ja johdolla, aloitettu standardien uudistamistyö.

Lisätietoja

PersID-hankkeen raportit ovat saatavissa hankkeen sivuilta, osoitteesta www.persid.org (<http://www.persid.org>). Kansalliskirjastosta URN-kysymyksiin vastaa mielellään tämän artikkelin kirjoittaja

Kirjoittajan yhteystiedot

*Esa-Pekka Keskitalo, tietojärjestelmäpäällikkö
Kansalliskirjasto / Kirjastoverkkopalvelut
PL 26, 00014 HELSINGIN YLIOPISTO
Sähköposti: esa-pekka.keskitalo (at) helsinki.fi*



Tietolinja

- 1/2011
 - Pääkirjoitus
 - Pilvistä Prahassa
 - Kuolema MARCille?
 - Avoin bibliografinen data
 - **PersID**
 - Julkaisuista tulospisteiksi
 - Uusi käyttöliittymä julkaisuarkistojen digitaalisiin aineistoihin
 - Uutisia ja ajankohtaista
- 2/2010
- 1/2010
- 2/2009
- 1/2009
- 2/2008
- 1/2008
- 2/2007
- 1/2007
- Arkisto
- Artikkelihakemisto
- Lehden tilaus

Kirjastoalan etusivulle